DOCUMENT RESUME

ED 069 161 FL 003 419

TITLE	Testing Some English Language Skills: Rationale Development and Description.
INSTITUTION	Toronto Board of Education (Ontario). Research Dept.
PUB DATE	Mar 69
NOTE	40p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Acculturation; Auditory Perception; Educational Experiments; English; *English (Second Language); *Evaluation Methods; Evaluation Techniques; Intonation; Language Fluency; *Language Skills; *Language Tests; Linguistic Competence; Phonology; *Test Construction; Vocabulary Development
IDENTIFIERS	New Canadian Study

ABSTRACT

This paper describes the development of materials to test English language skills and discusses some of the procedures and rationale which were employed in that development. The materials were designed for use in the New Canadian Study (1967-68), which was concerned primarily with the school success of students who learned English as a second language. This document discusses the theory behind the development of these testing instruments and details of the six-part instrument devised to test aspects of auditory perception and vocabulary. Each part is considered and described. A final chapter presents a preliminary evaluation of the materials. Shortcomings mentioned include a lack of coverage of grammatical structure, the language of the directions, and the idea of time limitations. Future use of the materials is also discussed. (Author/VM)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 069161

RESEARCH SERVICE

issued by the
Research Department

FILMED FROM BEST AVAILABLE COPY

PL005 417

HE BOARD OF EDUCATION FOR THE CITY OF TORONTO

ED 069161

TESTING SOME ENGLISH LANGUAGE SKILLS: RATIONALE DEVELOPMENT AND DESCRIPTION

March, 1969

FOREWORD

Mr. Sugunasiri developed and prepared the test materials. This report is based on his extensive notes; he had no opportunity to review this text.

E. N. WRIGHT.

CONTENTS

	Page No.
CHAPTER I -- INTRODUCTION	1
CHAPTER II -- LANGUAGE TESTING	6
What to Test	6
How to Test	10
Outcomes	12
CHAPTER III -- DEVELOPMENT OF THE ENGLISH COMPETENCE TEST	13
Parts I and II	13
Part III	20
Part IV	23
Part V	24
Part VI	25
CHAPTER IV -- A PRELIMINARY EVALUATION	27
Shortcomings	27
Future Use	31
Closing Exercises	32
REFERENCES	33
APPENDIX I	35
APPENDIX II	36

TESTING SOME ENGLISH LANGUAGE SKILLS: RATIONALE DEVELOPMENT AND DESCRIPTION

CHAPTER I -- INTRODUCTION

This paper describes the development of materials to test English language skills and provides some of the procedures and rationale which were employed. The materials were designed for use in the New Canadian Study (1967-68) which was concerned primarily with the school success of students who learned English as a second language. As a part of the Study it was desirable to use some instrument to assess the way in which English was understood and used by such students -- qualitatively and/or quantitatively. It soon became apparent that this would not be an easy task because no existing test seemed suitable for the purpose.

One early idea was to construct a form of vocabulary test which would explore richness and depth of vocabulary, not just in terms of the number of meanings an individual might be able to assign to a word (e.g., "strain") but also in terms of his ability to use and understand this multitude of meanings in context (e.g., "The quality of mercy is not strain'd...."). It is recognized that abstract words such as "quality" and "mercy" vary in meaning from culture to culture. Thus, to the native speaker of English who sees them in the above context, "quality" and "mercy" have a great number of loadings or referents peculiar not only to the way in which they are juxtaposed in this particular sentence but also peculiar to his cultural heritage. Perhaps specific reference might even be made in a discussion of this short sentence to North American or even Ontario usage, at least by the speaker to whom such usage is non-consciously natural. The

extent to which a New Canadian was aware of and could use these loadings would be a good indication of his acculturation as well as of his ability in English.

Such a test would likely be limited in usefulness, "unfair," and extraordinarily difficult to construct and validate. The population of the New Canadian Study was to include a representative sample of students chosen from Grades 5, 7 and 9 -- including students who had recently arrived in Canada, students who were born in Canada but whose parents spoke another language, and students who were English speaking monolinguals. Because of this variety, one immediate problem was vocabulary, i.e., the vocabulary of some of the students might be far below the level of vocabulary used in the test. Thus, no differences in degree could be ascertained among such students below a certain level. This problem arose not only in considering construction of a "language-in-culture" test, but also in considering other tests already in existence.

Another difficulty with a vocabulary test is that vocabulary and intelligence are closely related. While both above and below average children speak their native tongues fluently, the above average child usually has a more extensive vocabulary. Thus, a straight vocabulary test is one that tests intelligence as much as, or perhaps more than, anything else. In this part of the Study it was planned to try, as far as possible, to avoid testing intelligence.

English "Competence"

Such considerations led finally to the linguistic concept of "competence." To understand the concept of competence, it is first necessary to understand that a language is a finite system. Gleason (1961, p. 50) has counted 46 meaningful sounds (phonemes) in English. Paul Roberts (1956, pp. 298-299) notes

that there are seven basic (structural) patterns of English. Thus, it can be said that English is a finite system. A child reaches linguistic adulthood when he has mastered all the finite sets of linguistic elements. The child can then manipulate the known finite sets of linguistic elements to produce an infinite number of utterances that may never have been said earlier, by himself or anyone else. This is as true of English as of any other language. A native speaker (or for that matter anyone who has mastered English as a second language) can continue producing an infinite number of new utterances. This is the process of creative encoding. Other native speakers who hear such utterances accept them as utterances of their own language; this is creative decoding. This creative aspect of language, both productivity and receptivity, can be called the native speaker's competence.

There is no aspect of language which is not creative if we take an extended view of creativity. We cannot be certain whether a child's first speech is creative or imitative, but, as the child attains linguistic adulthood, he is certainly using his knowledge of his language, even when he produces utterances that he has heard. No two situations are identical and each situation, which by definition is new, demands novel patterns of linguistic communication where the language used can hardly be imitative. It is a ceaseless process of creating the "unknown" from known elements. Seen this way, language is always creative.

At this point a distinction must be drawn between competence and performance. At any given time-point the human organism has stored up a vast quantity of information about everything around him, including language. It would never be possible, even for the most insightful investigator, to find out exactly how much a person knows of a given subject area. A student's score of 100% in mathematics tells us only that he is capable of performing

all the mathematical calculations about which he has been questioned. It does not tell us how much more he is capable of doing. Nor does the score of less than 100% take into consideration other related factors such as memory or emotional mood which determine a person's behaviour just as much as knowledge does. Thus, it seems useful to distinguish two aspects of a person's linguistic behaviour: his total knowledge and the expression of that knowledge in a given situation. The former, Chomsky calls competence, the latter performance (Lyons and Wales, 1966, pp. 137-138). Since the number of utterances a person is capable of producing is infinite, but the actual number he produces at a given time is finite, performance falls far short of competence.

This introduction is being written after the materials were developed and used. Briefly it summarizes the goal that was set, an unobtainable goal. The initial posture during the planning stages was to accept no conceptual limitations but rather to devise instruments which could adequately sample a person's knowledge of the English language. A useful label for such a set of instruments was an "English Competence Test" (E.C.T.). The label remained long after it was apparent that the materials developed in a few months of concerted effort were tapping only some aspects of performance and were thus not even close to sampling competence.

The next section briefly presents some of the ideas and strategies which were considered during the early discussion stages. This is followed by a detailed description of procedures and rationale underlying the specific sub-tests. The concluding chapter discusses some of the limitations of this material.

As will be seen, the word "competence" should now be stricken from the title of these sub-tests. Following the analogy from mathematics,

above, it can be seen that a useful assessment of some language skills was made though not enough of them to infer an individual's level of competence.

CHAPTER II -- LANGUAGE TESTING

The following section discusses some of the principles underlying language testing. These will put in perspective some of the principles underlying the development of the English Competence Test (E.C.T.)

What to Test

The first question is "What do we test?" when assessing English as a second language. Dr. Finocchiaro, the distinguished American scholar of English as a second language, gives the answer:

> "...we should test everything we consider important enough in language learning to teach."
>
> (Finocchiaro, 1967, p. 111)

Components: Isolated or Integrated?

English, like any other language, is made up of a number of components. It is these that should be tested in a learner of English as a second language. But since speakers do not use these component parts in isolation, but rather integrate them in speaking, the question arises: do we test the component parts separately or in combination?

Lado has this to say on the subject:

> "...all we hope to find out in a foreign language test is how well he [i.e. the learner] has mastered the elements of the system of signals that constitutes that foreign language."
>
> (Lado, 1950, p. 77)

Elsewhere, in his book Language Testing Lado further clarifies:

> "...whether we test the elements of language separately or as part of one of these five skills [listening, reading, speaking, writing, and translation] we are still testing language. The decision between testing skills or testing elements...will be made on the basis of what information we need concerning the student's knowledge of the goal language."
>
> (Lado, 1961, p. 204)

In the present situation, a "...full-scale research project involving New Canadians..." had been requested by the Board. Although the term "New Canadian" is being defined in several ways, for purposes of analysis, it was for this part of the Study considered primarily in terms of language. The Background Information Questionnaire asked, "Before learning English did you learn any other language?" The subject who answered "yes" to this question would be considered a "New Canadian."

In this category, then, were children whose acquisition of English varied anywhere from zero (no English) to a hundred per cent (complete mastery). Because the information to be sought from the population concerned the extent to which English linguistic habits had been acquired, an instrument was required which could be used with most, if not all, New Canadians. A test of production (i.e. speaking) would not be usable with the population because of the existence of those subjects with zero or near zero ability in English. There would be other subjects in the population who spoke English, but poorly -- strongly accented, for example, or inadequately, as in the case of the person who says <I go>[1] instead of <I'm going>. An

---

1 At this point it is necessary to introduce the reader to certain symbols which are used throughout this paper. The written word "through" does not exactly correspond to its pronunciation: "ough" here is pronounced as in "boo." If the exact pronunciation is shown in writing (by using a special alphabet which uses mostly Roman letters) then we have a phonemic transcription, but when it is written with the regular spelling, then we have a graphemic transcription. In this paper, the graphemic transcription of a spoken word(s) will be written within wedge brackets (< >), as follows: <true>. The phonemic transcription will be written within slanted lines (/ /), as follows: /true/.

instrument of integrated skills would only tell us that he did not speak English well; it would not tell us how far he had proceeded along the path to learning English.

There was yet another problem to be considered. In the case of a baby learning to speak the mother tongue, it has been reported that "...there is apt to be a lag of 2 to 7 months from first learning to utterance" (Penfield and Roberts, 1966, p. 243). It is probable that in the case of a second language learner as well there would be a marked time lag between learning and utterance[2]. Thus, there is no way of knowing to what extent the utterance <I go> in the above example is an indicator of the speaker's knowledge of English. It may even be true of the New Canadian who speaks no English at all that he has "knowledge" of English but that the first break-through in the nerve paths leading to articulation have not taken place. This position is incidentally congruent with Chomsky's distinction between linguistic competence and performance -- that the latter, being behavioural, falls short of competence i.e. a person's knowledge.

Even if a test of production, or performance, had been deemed desirable, there was also the obvious, practical problem of testing production. It would indeed be an enormous, almost insurmountable task, to record and analyze the speech of some 6000 students who would constitute the study population.

Ideational Fluency

There was another possibility in addition to what has already been described -- a test of what might be called ideational fluency. Was there any way of testing the creative aspects of language usage? Ideational fluency must be an extension of linguistic competence; therefore why not include a test of ideational fluency?

---

2 See McNeil (1968) for a further discussion on this.

If a child's (age 4 or above) speech is compared with that of an adult, one of the differences to be found is in the richness of expression of ideas[3]. Because the adult knows more words, he is likely to express an idea better (e.g., more explicitly).

The New Canadian, since he already speaks his own language, certainly has some ideational fluency. For example, the neurological circuitry pertaining to the concept <the horse is running> is already present with reference to his own language. The problem for the New Canadian is to transfer that idea (concept) into the sound units of English. (If he lacks a concept then he must learn both the concept and the appropriate sound units of English.)

A suitable test of ideational fluency would at best tell us whether the student has knowledge of a particular concept, rather than whether he is able to express or hear it in English. In other words, the failure on the part of a New Canadian on a test of ideational fluency in English might be due to either his lack of English competence, or lack of ideational competence, or both. We would not know which. If he were tested in his own language, there would be no direct indication of English competence.

Reading and Writing

Writing is not synonymous with language: it is only a representation of language by means of symbols. Writing is one step removed from oral language. It is essentially a matter of tracing symbols representing the sounds of language. But to trace them, one first must have the sound units. Thus, writing is an extension of language; or writing is a secondary (not in importance, but developmentally) language ability. So is reading. Before

---

3 Compare in this connection a creative writer and an ordinary adult. Since the creative artist handles language with more dexterity, there would be more richness of expression in his writing than that of the non-writer.

one can read (and understand) one must "know" the language. If writing is tracing symbols, reading is meaningfully vocalizing the symbols or units of symbols, representing sound.

The man who reads English knows his Roman alphabet. French is also written in the same Roman alphabet, but the English-reading man perhaps cannot read French. He could pronounce each sound, or combinations of sounds (i.e. words) without understanding any of what he was vocalizing (or seeing). Similarly there is the case of the New Canadian child who "reads" English out loud from the page but is unable to tell what he has read. Both cases illustrate that reading, like writing, is different from oral language.

How to Test

Basing Test on 500 Words of Highest Frequency

A further limitation was determined by the nature of the study population but had a theoretical basis as well.

Vocabulary is a component of language that continues to enlarge. The notion that language is synonymous with vocabulary is therefore a misconception. (A Canadian child with a very low vocabulary can indeed speak English as fluently as an adult, if fluency is defined only as the ability to manipulate the components of the language.) To limit the influence of the vocabulary component on the sub-tests it was decided that the materials should be constructed (as far as possible) using only the Lorge-Thorndike List of the most frequently used 500 words.

In this list are contained nearly all the functors of English, i.e. those words that have primarily a grammatical, or functional, meaning. To understand this concept, take the word <is> in the utterance <Mom is going>.

The only way we can give it meaning is by referring to its function in an utterance. Thus <is> denotes the present time of an action with reference to one doer. A grammarian would say that it denotes the present tense singular. Further we can say that it occurs before the doer in the statement, but again this is contextual. So we say <is> has only a functional meaning, and is therefore a functor[4]. Words with non-grammatical meaning, e.g., <mom> are called contentives.

An analysis of the language used by a child aged four who is said to be a linguistic adult but whose vocabulary is nowhere near that of an average adult would show that the proportion of functors used by him is higher than the proportion of contentives used. Partly this is because the functors are quite limited in number -- Finocchiaro (1964, p. 27) counts 154, whereas the number of contentives is many times larger. It stands to reason that the New Canadian learning English as a second language will have to master the functors in order to be able to manipulate the language (i.e. be fluent in terms of the limited definition): as vocabulary expands the number of contentives increases. To have more verbal labels, however, does not ensure an increased ability to manipulate them, use them in sentences, phrases, or in the past and present tenses. Therefore it seems most appropriate to use the word list for the purposes which have been outlined.

Furthermore, the contentives included in the first 500 words were basic enough to be used with New Canadians having limited exposure to English, thus extending the usability of the E.C.T.

---

4 Hockett (1958, p. 264) prefers functors and contentives to function words and context words, since both categories include forms that are not words, e.g., <-ing> (functor). He lists four types of functors:

1. substitutes;	examples:	<all>, <it>, <the>
2. markers;	examples:	<and>, <in>
3. inflectional affixes;	examples:	<-s> plural, <-ed> past tense
4. derivational affixes;	examples:	<-y> e.g., milky; <-ness> e.g., happiness.

There can be two criticisms against using the Lorge-Thorndike list as a basis for a discussion on language: its basis is the printed word and not children's speech; the count was taken 24 years ago, and as such is not representative of today's usage.

The first objection can be overruled by saying that printed material may not be far removed from standard usage, and certainly not to such an extent as to be non-representative of speech. Three reasons can be advanced on the second count: (a) the first 500 words of any given time would likely include all the functors; (b) there would probably be a large enough number of words in the list that are representative of the current usage; (c) time is more likely to increase the usage of certain contentives, e.g., "atomic" than functors.

Using the first 500 words is also justified as New Canadians' ideational fluency, or reading and writing ability, was not being tested.

Outcomes

An early decision was made not to use standardized reading tests. In addition a picture vocabulary test was selected as a component of the test battery. It appeared that any material needed to assess English language skills might best be built. Although such instruments would provide only coarse scales they would shift the focus from classroom performance to language performance. Rapidly an attempt was made to define operationally some of the component English language skills and construct items that would require these skills for a correct answer. The following chapter details the steps in constructing these sub-tests.

CHAPTER III -- DEVELOPMENT OF THE ENGLISH COMPETENCE TEST

This chapter uses the label which by now had stuck as a term to describe the test materials (indeed even the answer sheet was labelled "English Competence"). The fact that this label can now be clearly seen as inappropriate hopefully will not turn the reader away from a consideration of the details regarding content selection and item development.

Format of the Test

The test consisted of six parts, divisible into two broad categories: auditory, and paper and pencil.

Parts I to III tested aspects of auditory perception, and were pre-recorded on tape. The directions to the students, the practice examples, and the test items themselves were also taped in English. Parts IV to VI tested vocabulary aspects of language and were in printed form. All answers were recorded on a separate Digitek answer sheet. Students also had a practice sheet to use with the instructions for each part of the test.

The total running time of the taped materials, including directions and test items, was 36 minutes. The time limits on the remaining parts were 5, 3, and 4 minutes respectively, exclusive of time for directions.

Parts I and II

The purpose of Parts I and II was to test two basic elements of language: sound discrimination and sound recognition. Given two (or more) speech sounds or sequences of sound, a competent speaker is able to identify whether the two sounds are the same or different. This is discrimination. Sound recognition refers to the ability of a speaker to identify a speech sound or sequence of sounds, using an existing referrent which will probably

include information about both sound and meaning. Because different sounds and combinations of sounds are significant in different languages, recognition and discrimination of the sounds in English are an important aspect of English competence. Probably in practical terms there is little difference in difficulty between recognition and discrimination.

Part I consisted of 45 minimal pairs. A minimal pair is two words that sound alike except for one phonemic difference, e.g., <pit:bit>. The two stimulus words were said with an interval of approximately one second between them. All items were presented by the same voice (female adult). The subject was given 4 seconds to decide whether the two words were the same or different, and to mark his answer by darkening S (for same) or D (for different) against the item on the Answer Sheet.

Example 1 -- coat	coat	S▯	D▯
Example 2 -- bag	back	S▯	D▯

Part II also included 45 items and tested sound recognition. The voice (adult female) first presented one word. Then, approximately 2 seconds later, two more words were presented by the same voice, one of which is the same as the first word uttered. The subject was to indicate, by darkening A or B on the Answer Sheet against the particular item, which of the latter two words is the same as the first[5].

The same voice used in Part I presented all 45 items.

Example 1 -- poor	power	poor	A▯	B▯
Example 2 -- tick	tick	Dick	A▯	B▯

---

5 As a response technique, Lado (1961, p. 54) favours "...marking the choices that are the same rather than those that are different because it approaches slightly closer the operation of language. The listener identifies what a thing is first, and then is able to tell what it is not."

Basis of Item Selection for Parts I and II

Gleason's list (1961, p. 50) of 24 English consonants[6] and nine vowels served as the basis for the test items.

Minimal pairs were then sought based on the following criteria:

I - Consonants
1. voiced/voiceless contrast
   example: /p/: /b/
2. contrast in initial and final sounds
   example (a) pin:bin
   (b) lap:lab
3. contrast in points of articulation, in nasal consonants only[7]
   example: /m/: /n/

II - Vowels
1. front/back dimension[8]
   example /i/: /u/

---

6 See Appendix I for the list.

7 Consonants are measurable or describable on three criteria: voiced/voiceless, point of tongue contact, and degree of constriction (i.e. whether lung air is completely checked and then released -/p//b/ etc., - or is partially constricted -/f//s/ etc.). All nasal consonants are voiced and produced with total constriction, so that the only contrast is with reference to the point of articulation.

8 "Three variables in the positions of the vocal organs are particularly significant in the phonetic description of English vowels. The most important is the position of the highest part of the tongue. This varies in two dimensions. It may be relatively high, mid, or low. It may also be relatively front, central, or back. Note the difference in meaning between central (intermediate between front and back) and mid (intermediate between high and low). These two variables provide symmetrical charting of the English vowel phonemes:

	FRONT	CENTRAL	BACK
HIGH	i	ɨ*	u
MID	e	ə	o
LOW	æ	a	ɔ

The third variable which is of importance in English vowels is the rounding of the lips. In /u/ there is always moderate rounding. In /o/ there is usually somewhat less, but always enough to be noticeable. In /ɔ/ the rounding is still weaker. In some pronunciations it may be so slight as to be hardly noticeable, or even lacking. The front and central vowels are never rounded." (Gleason, 1961, pp. 35-36) *Not used in test materials as native speakers maintained that it does not occur in Canadian English.

2. high/low dimension

   example /i/: /e/

3. initial, medial and final

   example (a) it:eat

   (b) bat:but

   (c) way:we

4. length

   example (a) bit:beat

   (b) coat:caught

5. nuclei (glides)/nuclei[9]

   example bite:bout

6. nuclei/single vowel

   example but.bout

III - Other Contrasts: There are other significant contrasts which would not be covered by the above. An attempt was made to include these, too, albeit not comprehensively[10].

example: (a) lamp: ramp[11]

(b) art: heart[12]

thing: think

veal: wheel

---

9 Vowel + semi-vowel = nuclei. See Gleason 1961 (pp. 34-35) for a list.

10 Clusters (e.g., str- as in "street") should also have been included in the list, but have not been. This is a shortcoming.

11 Example from Perren (1967, p. 24).

12 These examples came from the author's personal experience in teaching English as a second language.

Once meaningful[13] minimal pairs were found to satisfy the above conditions[14], three pairs of words were made for each pair. Two pairs consisted of one member of the minimal pair repeated, and the third was the minimal pair itself. Example: pin : pin

bin : bin

pin : bin

The order was arbitrary. Again, quite arbitrarily, the two members of the minimal pair were almost always written in the following order:

voiceless : voiced, e.g., /p : b/

front : back, e.g., /m : ŋ/ ; /i : u/

short V : long V, e.g., /i : i/

Generally, minimal pairs for the consonants were placed first in the list, to be followed by those for vowels. Other elements tested were placed as they were created.

The next step was to eliminate those pairs which included one member used on more than one occasion. For example <bit> was contrasted with <bet> and with <bat> again. Thus, there would have been two minimal pairs which included <bit>.

This process of elimination left 94 pairs out of an earlier 102. The first 90 pairs were then renumbered using a table of random numbers. This randomization helped to overcome any bias that might have occurred by placing the pairs for the consonants first, followed by those for the vowels.

The first 45 of these pairs constituted Part I of the test, and the next 45, Part II.

---

13 "Meaningful" here means "those that are acceptable as English word(s)." However, in the final test there were two exceptions: bon, con.

14 One pair (thaw : though) had two differences: voiceless:voiced and O:ɔ (vowel sound as in <caught> for want of a better pair.

A further step in Part I was to replace a word occurring in an earlier pair, with a word having the most number of parallel features, but still testing the same element. For example <vole> occurred with <foal> in pair number 30. Pair number 45 was <vole : vole>, testing the /v/ sound. Thus, <vole> was replaced by <vote> and pair number 45 became <vote/vote>. This step was taken to avoid any learning effect.

The format of Part II (example: cab - cab/cap) did not allow pairs of similar words. Hence, the list had to undergo the following modifications before the final list was arrived at:

1(a) Wherever a pair of similar words occurred, the phoneme tested in that item was determined by reference to the minimal pair in which the word originally occurred, and the similar pair then replaced by the minimal pair.

(b) As was done in Part I, if any of the members of the minimal pair had been used earlier, the word(s) was/were replaced by a pair testing the same element.

One example of this process (1(a) and (b)) was item number 46 which originally was lab/lab. This was replaced by the minimal pair lap/lab. Since this minimal pair occurred in Part I, it was replaced by cap/cab. Sometimes this process resulted in retaining only the phoneme to be tested. Item number 89 thigh/thy was modified until it became thistle/this'll. This, incidentally, was the only example of using more than one word as a member of a minimal pair.

2. Once all the minimal pairs were thus revised and replaced, one member of the pair was randomly chosen as the first stimulus.

Discussion of Parts I and II

The basis for establishing test items for Part I and Part II can be questioned. The reason for deliberately selecting the test items first and then

randomizing and including contrasts that were not in the original list was to ensure that at least some significant contrasts like l : r (lest : rest) (Perron, 1967); d : ð (bade : bathe) (Lado, 1961, p. 52) were included. Such contrasts, though they pose no problems to the native speaker, give much difficulty to a second language learner, the difficulty depending on the sound system of his language.

Parts I and II had 45 items each, as opposed to less than half that number in each of the other parts, because necessary pilot testing and item analysis had not yet been done. A second reason, the more important one theoretically, came from the nature of the study population, which was composed of at least 30 ethnic groups. While a particular sound contrast might prove difficult to one ethnic group, another group might find a different contrast to be more problematic. A large list is more likely to include more representative problems than a short one.

Technique

Lado (1961, pp. 53-54) in his book "Language Testing" says the following of the technique used above:

> "This is an excellent technique for classroom testing. For formal tests, it suffers from the problem of all two-choice tests: the effect of wild guessing is quite heavy... as for validity, it is more valid than even extended observation of students in their everyday use of the foreign language...In the ordinary use of the language it is difficult to know if the student has perceived the sound contrast, guessed at the meaning, or understood from the context[15] rather than through the words containing the difficult sounds. In this technique, on the other hand, we are able to control the non-language factors and the language factors that are primarily pronunciation."
>
> (Lado, 1961, pp. 53-54)

---

15 If we were testing the integrated skill rather than the elements of language separately, this, of course, would not be an objection.

Part III

Part III was comprised of 24 items designed to test another aspect of auditory discrimination -- intonation. In English the same string of words can be said in different ways to mean different things. The purpose here was to see whether the subjects could discriminate these differences on hearing a given utterance.

The material was presented on tape. The subject first heard an utterance (primary stimulus). After a few seconds, he heard two more utterances (secondary stimulus), which gave two possible interpretations of what he first heard. He was to indicate the correct choice by darkening A or B; example:

I want to go home. A.[] Home is where I want to go. B.[] It's I who want to go home, not him.

Format

In addition to hearing the two choices, the subject also saw them on paper. This was to ensure that the subjects were not being tested on the secondary stimulus rather than on the primary stimulus. The point of this was to determine whether the subject could hear the primary stimulus correctly. Thus, it was deemed necessary to give as many cues as possible to the subject to minimize his difficulty in understanding the choices. (Providing the alternatives in a students' mother tongue was not practical nor was it known how well the concepts could be translated.)

The decision to let the choices be both heard and read was primarily determined by the analysis of results of pilot runs with grade five students. Choices were presented to subjects in three different ways. One group only heard them, another only read them, and a third group both read and heard them. A comparison of means showed a mild (not statistically significant) difference among the three treatments in favour of "read and hear," when the

overall groups, made up of both native and New Canadian students were considered. However, there was no difference when only the native Canadians were compared. The differences were attributed to the New Canadians.

At the theoretical level, Perren (1967, p. 24) feels that a subject is "...listening with a purpose..." when he "...sees the choice of response before or simultaneously hearing the statement." Ausubel (1964, p. 423) agrees that "Simultaneous reading support can furnish the necessary cues for meaning...."

Admittedly, seeing the choice at the same time as it is heard puts the good reader at an advantage, both over the poor reader and the non-reader. However, it is likely that even a poor reader can pick up certain cues, especially if he is also hearing the words. Such a person may get information from either or both hearing and seeing. This would put only the non-reader at a great disadvantage. It was assumed that the study population would have a large percentage of readers, both good and bad, as opposed to non-readers.

The only other practical way to ensure maximum cueing was by repeating the choices: it was observed during the pilot runs with this treatment that the subjects were very restive and bored, since there was too much idle time, particularly when both the primary stimulus and the secondary stimulus were repeated. A repetition of only the latter would have led to cognitive dissonance, since by the time the subject finished hearing the choices the second time, the primary stimulus would no longer be "ringing in their ears."[16] Furthermore, repetition would have increased the artificiality of the linguistic situation.

---

16 See Hutchinson in Valdman (1966, p. 225) for a discussion.

Two-Choice vs. Multiple Choice

The only justification for having a two-choice rather than a multiple choice test was the difficulty of getting more than one good distractor. Thus, Part III suffers from the problems of all two-choice tests, specifically the fact that the subject should be able to score 50% by guess-work alone.

Voice

Unlike Parts I and II, this part of the test was recorded in different voices, to get the effect of Standard Variety.

> "Standard Variety means any variety spoken by educated speakers of that language from any of the regions where it is spoken natively."
>
> (Lado, 1961, p. 47)

A Toronto student is bound to hear a variety of English spoken by his teachers and peers. To approximate the Standard Variety, five voices from different parts of Canada, one adult female and four adult males, were used to record the primary stimulus. Voices were randomly assigned to each test item.

The secondary stimulus (the two choices for each item), however, was consistently presented in the same adult male voice.

Basis of Item Selection

Gleason (1961, Chapter 4), Hockett (1958, Chapter 4), and Long (1961, Chapter 20) served as the basis for the items in Part III. The initial list consisted of 33 items, nine of which were deleted from the final list as being non-discriminatory, on the basis of pilot runs. These were primarily items which 50% or more of the high scoring students got wrong or items on which the low scoring students did as well as the high scoring students.

The items finally selected were placed in order of difficulty, in an attempt to lessen anxiety that might be associated with this novel test procedure.

Part IV

Both Parts IV and V tested vocabulary. Here again the words were limited to the first 500 words of the Lorge-Thorndike count. Part IV was a test on contentives and included six words each of which have more than one meaning. The testees were given twelve items which gave meanings of the six words.

	A	B
Example --	1. part of a play	A. head
	2. a bill presented in Parliament	B. act
	3. a part of the body above the neck	
	4. to do something	

Format

This part was a paper and pencil test, and took the form of two lists, A and B. List B gave the test words, and list A the meanings. This reversed order was determined by the requirements of the optical scanning machine used to score the answer sheets. A change in the order would have resulted in the answer sheet having twelve choices. The subject thus first read the meaning, and then found the word that fits the meaning.

Out of the six test items, two had three meanings each, two others had two meanings each, and the other two had only one meaning each.

Basis for Selection

Since the interest was in words which had multiple meanings (and there were not too many in the list of 500 words) initially only seven items were picked. The Concise Oxford Dictionary provided a list of 21 meanings

that could be expressed using the basic 500 words. Items in both A and B which, on the basis of pilot runs were either too easy or too difficult, were dropped. The final test therefore, had six test items and twelve choices for meaning.

Part V

The difference between Part V and Part IV was that although both test vocabulary, Part V tested functors. Functors are not only one of the basic features of grammatical signalling in modern English, but pose particularly difficult problems for both the learner of English as a second language and the child from a low socio-economic level. Schutz and Keisler (1968, p. 213) report that their results "...support the findings of other investigators that lower class children are most deficient with functors" (as compared with nouns and verbs).

The subjects were given four functors and twelve choices, each of which was a complete sentence when the blank was filled in with a functor.

Example -- It's time _____ bed.
A. for
B. after
C. out
D. along

Format

The format was similar to Part IV. One functor was the correct answer four times; two functors were the correct choice three times; and, the other two were used twice.

Basis of Item Selection

The criteria determining the selection of items in this section were, again: (a) test words were to be from the Lorge-Thorndike list;

(b) sentences in which the words occurred were also to be made up of words from the list.

An attempt (based on common knowledge) was also made to ensure that the concepts in the choice items were as "common" as possible.

These criteria initially provided a list of six test items and 22 choices, which was reduced to four test items and twelve choices, after elimination of the easiest and the most difficult ones on the basis of pilot runs.

Part VI

Part VI consisted of ten items designed to test idiomatic usage. Idioms are different from vocabulary items in that the combination of a particular sequence of words carries a meaning different from the meaning of the individual words that make up the idiom. English employs this technique of idiomatic usage heavily, and therefore a grasp of this element is an important skill.

In this section, which was also paper and pencil, possible meanings of each idiom were provided. The rigorous requirement of limiting the vocabulary in the choices to the Lorge-Thorndike list had to be given up here for stylistic as well as other reasons. It was simply not possible to make up four choices for ten items within the range of 500 words. However, an attempt was made to keep "outside" words at a minimum. Again, the choice was indicated by darkening A, B, C or D.

Example - What is the meaning of ...

1. I haven't seen him for ages.

A. I haven't seen him for some time.

B. I haven't seen him since he was four.

C. I haven't seen him for a few hours.

D. I haven't seen him since the meeting at which he spoke for ages.

Basis for Item Selection

Test items were invented as ideas occurred since a good list of idioms could not be located. Whatever the biases in this process, in the pilot runs, ten of the twelve items originally selected were found to be discriminatory, and were retained in the test. Again, the items were placed in order of difficulty. The four answer choices for each item were randomly ordered.

CHAPTER IV -- A PRELIMINARY EVALUATION

Within six months of beginning to look for a way of assessing a person's use of English, the materials, in their present form, were prepared for administration to an expected student population of 6,000. The shortage of time is not an excuse, but it limited the amount of literature that could be reviewed in advance and the amount of material that could be developed. Pilot work with materials was minimal and item analyses will wait until it is determined whether there is any continuing need for the test materials.

Shortcomings

Where's English Structure?

Grammatical structure constitutes an essential sub-system of language. Any language test which does not have a section requiring this knowledge should be considered thoroughly unsatisfactory. On these grounds alone the materials must be deemed incomplete, for they do not have items on structure. Strategies for such a section had been outlined; however, there was not time to develop these strategies.

"I don't understand your directions."

One area requiring further research deals with the language in which directions were given. If the purpose is to test the student's knowledge of English, it must be done with carefully selected test items and not with standard directions. Otherwise, it becomes an exercise in understanding directions. A child whose English competence is poor would be unable to understand instructions in English. A solution is to give instructions in his mother tongue.

In the present study with over 5,000 students whose first languages represented eighteen or more linguistic groups, such a procedure was simply out of the question. As the test has tape recorded instructions there is a possibility in the future of preparing separate tapes with the instructions in different languages. This area (giving directions in a language other than the language of the test) has not, to the best of this writer's knowledge, drawn the attention of researchers. The effe ts of giving directions in the subject's mother tongue requires research. Certainly any test translations ought to be standardized. The onus of translation should not be on the individual school that administers the test. It must come in the package.

Informal feedback from the schools indicated that the taped instructions were difficult. As a result of pilot work, the taped instructions were completely revised twice. The major difficulty is not yet known; however, the medium itself is suspect. Taped tests are so novel that it is quite conceivable that they interfere with the mental set for test taking which North American students have developed. Actually, the major complaint was that the taped material took far longer to give than the actual playing time of the tapes would suggest. Preliminary examination of a handful of answer sheets suggested that once the students were working with the test material from the tape they had relatively few problems.

The Time-clock is North American!

North American culture is a strongly time-oriented one. The Canadian-born child who has been reared in such a culture finds no problem -- or at least no surprise -- in facing a test which has a time limit. However, the New Canadian subjects of our study may not have been necessarily time-oriented, particularly if they came from a non-industrial

background. Under such circumstances a speed test may have evoked anxiety in the subject which could seriously impede his test performance. (Students from certain Canadian sub-cultures are also peralized by tests that insist on speed.)

Thus, it is suggested that the E.C.T. be made a "power-test," in which subjects are given as much time as they want to answer. Of course, nothing in the content need be changed for this purpose, only no time should be stipulated. This could be done if the test were administered on an individual basis.

The effect of timing on performance also requires study. The question is particularly relevant to the recorded part of the test: whatever the cultural background, it may be best to respond while the stimulus is still ringing in the ear.

How Valid is the Word List?

The entire E.C.T., with the exception of a few items, was based on the Lorge-Thorndike list of the first 500 words (see pp. 11 - 12). A better list of words would be one based on a representative sample of native Canadian fifth-graders, drawn from varied ethnic and socio-cultural groups. Fifth-graders are suggested because Grade 5 was the lowest grade level in the study. To make the E.C.T. applicable to lower grades that particular age group should be sampled to ensure that the words used are within the students' grasp.

Another advantage in working from a new word list would be that this would permit the use of the first thousand words, without seriously undermining understanding. With the Lorge-Thorndike List, it was with less certainty that additional words could be chosen from the second 500 words, since the list is outdated. The use of 1,000 words would be a valuable advantage that would enable the secondary stimulus (particularly in Parts III and VI) to be worded less awkwardly.

Words Not in First 500 of List. The original intention of restricting every word in the materials to the first 500 words could not be implemented completely because of problems in wording. Thus, in Parts III and VI, the secondary stimulus was not always restricted to the list, although an attempt was made to ensure that such was the case in Parts IV and V.

It must be reported that a few times words from the list of 501 - 1000 words were used and even a few other words. Some of these words are given below:

doubt, hen, chipmunk, cat, cupboard, spent, Communist;

un-, -ing, -ed were some other units that were used with words from the list. The words outside the list were considered to be familiar enough in the day-to-day life of the native Canadian subject.

Your Bias is Showing!

A final shortcoming was that test items were deliberately and specifically chosen and thus may be biased in some way. Although the instruments' validity can be questioned on these grounds, such selectivity was necessary. Lists do not exist from which the material for the particular component parts could be randomly drawn. Such a random selection would not ensure the inclusion of all important linguistic problems. The selection still does not adequately recognize the view that certain linguistic sounds and combinations carry more meaning and functional load than others.

One particular problem of item selection in language tests was pointed out by Lado (1950, p. 79):

> "Trying tests on native speakers of a language will show some of the things that should be left out of the tests, but it does not tell us what we should put into them."

Difficulties in the selection of phonemic contrasts are shown by Perren (1967. p. 25):

> "...data is available in the relative frequency of occurrence of phonemics in particular varieties of English, but these provide no certain indication of the relative functional load which they carry[17] ...selection of significant contrasts become particularly difficult when test...has to be designed for application to students with varying mother tongues."

Future Use

Toronto has an extensive programme for New Canadians of instruction in English as a second language. To the best of this writer's knowledge, however, there is no (standard or non-standard) test used to determine the students' level of competence in English. Probably none suits the schools' requirements since most language tests are designed for students who already have a fairly high level of English competence.

One value of the E.C.T. lies in its potential use with students who arrive with a little knowledge of English. Since half of the Test was auditory, even those students with no writing or reading ability in English could be tested. Further, the Test was constructed on a sliding scale of difficulty. The New Canadian who is able to answer only Parts I and II (phonemic contrasts) possesses less English competence than one who can answer Part III (intonation). A student who correctly answers Part VI (idioms) is certainly more advanced than the one who stops at Parts IV and V (vocabulary). It should be reiterated here that if such use of the instrument were to be made, the directions should be given in the native language of the student.

---

17 Data are available on the functional load of the commonest words. (See, for example, Lorge's (1949) semantic count.)

The test may eventually be useful for several purposes, including programme placement, diagnosis and even possibly as a measure of achievement. These possible applications would require normative data and also modifications and extension of the materials.

Closing Exercises

The E.C.T. is far from perfection. It was an attempt to solve some of the problems of testing English as a second language. The major contribution, it is felt, was the break from the traditional patterns used in current language tests. The vocabulary was limited to the most frequently used words. The testing situation could be managed by a student with a limited number of English words at his command. It is hoped that these tests of language will be uncorrelated with intelligence.

A second contribution may be the shift of emphasis to testing language skills without reference to the classroom. Answers about the value of the test and further relevant data will follow. An analysis of the results of the E.C.T. as administered to the more than 5,000 students as part of the New Canadian Study is planned as a separate report[18].

---

18 The materials used in the E.C.T. will be made available by the Research Department on special request.

REFERENCES

Ausubel, D. P. Adults versus children in second language learning: psychological consideration. MLJ, November, 1964, 48 (7), pp. 420-424.

Chomsky, N. Syntactic structures. The Hague: Moulton, 1957.

Chomsky, N. Language and the mind. Psychology Today, February, 1968, pp. 48-51, 66-68.

Finocchiaro, Mary. English as a second language: from theory to practice. New York: Regents Publishing, 1967.

Gleason, H. A., Jr. An introduction to descriptive linguistics. New York: Holt, Rinehart & Winston, 1961 (revised edition).

Hockett, C. F. A course in modern linguistics. New York: Macmillan, 1958.

Lado, R. Linguistic science and language tests. Language Learning, 1950, 3, pp. 75-82.

Lado, R. Language testing. London: Longmans, 1961.

Long, R. B. The sentence and its parts: a grammar of contemporary English. Chicago: The University of Chicago Press, 1961.

Lorge, I. The semantic count of the 500 commonest English words. New York: Teachers' College, Columbia University, 1949.

Lyons, J., & Wales, R. J. (Eds.) Psycholinguistics papers: the proceedings of the 1966 Edinburgh conference. Edinburgh: Edinburgh University Press, 1966.

McNeil, D. Production and perception: the view from language. Ontario Journal of Educational Research, Spring, 1968, 10 (3), pp. 181-185.

Penfield, W., & Roberts, Lamar. Speech and brain-mechanisms. New York: Atheneum, 1966.

Perren, G. Testing ability in English as a second language: 3. Spoken language. English Language Teaching, October, 1967, 22 (1), pp. 22-27.

Roberts, P. Patterns of English. New York: Harcourt, Brace and World, 1956.

Schutz, S. R., & Keislar, E. R. Immediate memory of three word classes as a function of social class and grade. Paper read at the American Educational Research Association Annual Meeting, Chicago, February, 1968.

Valdman, A. A. Trends in language learning. New York: McGraw-Hill, 1966.

APPENDICES

APPENDIX I

List of English Consonant Phonemes as Given by Gleason (1961, p. 50)

(See Footnote 6)

/p/	as the underlined sound in	<p>in>
/b/		<b>in>
/t/		<t>in>
/d/		<d>in>
/k/		<k>in> : <c>at>
/g/		<g>ong>
/č/		<ch>eck>
/ǰ/		<j>eep>
/f/		<f>in>
/v/		<v>ictory>
/θ/		<th>in>
/ð/		<th>is>
/s/		<s>in>
/z/		<z>one>
/š/		<sh>un>
/ž/		<vis>ion>
/m/		<m>an>
/n/		<ma>n>
/ŋ/		<sing>
/l/		<l>ive>
/r/		<r>ing>
/w/		<w>in>
/y/		<y>oung>
/h/		<h>at>

Although most of the examples above have been in initial position, phonemes occur in initial, medial and final positions, with a few exceptions. For example, /ŋ/ doesn't occur initially.

APPENDIX II

Recommended References

A. Books

The following three books on Linguistics have been written for the layman. These books present the material in a non-technical and attractive manner:

Hall, R. A., Jr. Linguistics and your language. (Revised edition of Leave your language alone.) New York: Anchor Books, Doubleday & Co., 1960.

Sapir, Edward. Language. New York: Harcourt, Brace & World, 1921.

Scargill, M. H., & Penner, P. G. (Editors) Looking at language. Toronto: W. J. Gage, Ltd., 1966.

B. Periodicals and Journals*

English - A New Language. Commonwealth Office of Education, Sydney, N.S.W., Australia.

English Language Teaching.** The British Council, 65 Davis Street, London, England.

English Teaching Abstracts. The British Council, 65 Davis Street, London, England.

English Teaching News. The British Council, 65 Davis Street, London, England.

Language Learning: A Journal of Applied Linguistics 1522 Rackham Building, Ann Arbor, Michigan.

Linguistic Reporter. Center for Applied Linguistics, 1755 Massachusetts Avenue, Washington, D.C.

M. L. Abstracts. Fullerton, California.

The Modern Language Journal. Curtis Reed Plaza, Menasha, Wisconsin.

N.A.F.S.A. Newsletter. National Association of Foreign Students Advisers, 500 Riverside Drive, New York 22, New York.

---

* List from Finocchiaro, 1964, pp. 140-141.

** Not in Finocchiaro's list.